SUB C17

WHAT IS CLAIMED IS:

1. A method of determining a relationship between first and second textual inputs, the method comprising:

obtaining a set of relations;

identifying constituents in the first textual input that have the relations; and

determining the relationship between the first and second textual inputs based on the constituents identified.

2. The method of claim 1 wherein the determining step includes:

determining the relationship between the first and second textual inputs based on the relations.

3. The method of claim 1 wherein obtaining a set of relations comprises:

obtaining a hierarchy of grammatical relations; and

obtaining a hierarchy threshold based on a usefulness of grammatical relations in the hierarchy in determining the relationship between the first and second textual inputs.

4. The method of claim 3 wherein the determining step comprises:

determining the usefulness of identified constituents by locating the grammatical relations associated with the identified constituents in the hierarchy.

5. The method of claim 4 wherein the identifying step comprises:

identifying low ranked constituents having corresponding grammatical relations located in the hierarchy below the hierarchy threshold.

6. The method of claim 5 wherein the determining step comprises:

determining the relationship based on constituents in the first textual input, other than the low ranked constituents.

7. The method of claim 5 wherein the identifying step includes:

identifying high ranked constituents having a corresponding grammatical relation located in the hierarchy at least as high as the hierarchy threshold; and

annotating the high-ranked constituents with a weighting value which weights the high-ranked constituents higher than low-ranked constituents.

8. The method of claim 7 wherein identifying constituents in the first textual input comprises:

annotating the high-ranked and low-ranked constituents with fine values based on a location of grammatical relations corresponding to each of the constituents in the hierarchy, the fine values being indicative of relative usefulness of the

constituents in determining the relationship.

9. The method of claim 8 wherein the step of determining the relationship based on the constituents comprises:

determining the relationship based on the fine values associated with constituents.

10. The method of claim 4 wherein the determining step comprises:

preferentially matching terms in the first textual input against higher constituents in the second textual input having corresponding grammatical relations located relatively higher on the hierarchy than grammatical relations corresponding to lower constituents.

11. The method of claim 10 wherein the first textual input comprises a document and the second textual input comprises an information retrieval query and wherein preferentially matching comprises:

obtaining an index having entries corresponding to the document, the entries corresponding to only the higher constituents as opposed to the lower constituents; and

matching search terms in the query against the entries in the index.

12. The method of claim 10 wherein the first textual input comprises a document and the second textual input comprises an information retrieval query and

wherein preferentially matching comprises:

obtaining an index having entries corresponding to the document, the entries corresponding to the higher constituents having higher weighting values associated therewith and the entries corresponding to the lower constituents having lower weighting values associated therewith; and

matching search terms in the query against the entries in the index based on the higher and lower weighting values.

13. The method of claim 1 wherein obtaining relations comprises:

obtaining a hierarchy of case information; and

obtaining a hierarchy threshold based on the usefulness of a constituent having that case.

14. The method of claim 13 wherein the determining step comprises:

determining the usefulness of the identified constituents by locating the case information associated with the identified constituents in the hierarchy.

15. The method of claim 14 wherein the identifying step comprises:

identifying low ranked constituents having the case indicated by the case information.

16. The method of claim 15 wherein determining the relationship between the first and second textual

inputs comprises:

determining the relationship based on constituents in the first textual input, other than the low ranked constituents.

17. The method of claim 15 wherein the identifying step includes:

identifying high ranked constituents having corresponding case information located in the hierarchy at least as high as the hierarchy threshold; and

annotating the lower ranked constituents with a weighting value which weights the low ranked constituents lower than the high ranked constituents.

18. The method of claim 17 wherein identifying constituents in the first textual input comprises:

annotating the low ranked and high ranked constituents with fine values based on a location of the case information associated with each of the low ranked and high ranked constituents in the hierarchy, the fine values being indicative of relative usefulness of the constituents.

19. The method of claim 18 wherein the step of determining the relationship based on the constituents comprises:

determining the relationship based on the fine values associated with the constituents.

20. The method of claim 14 wherein the determining step comprises:

preferentially matching terms in the first textual input against higher ranked constituents in the second textual input having corresponding grammatical relations located relatively higher on the hierarchy than grammatical relations corresponding to lower constituents.

21. The method of claim 20 wherein the first textual input comprises a document and the second textual input comprises an information retrieval query and wherein preferentially matching comprises:

obtaining an index having entries corresponding to the document, the entries corresponding to only the higher ranked constituents as opposed to the lower ranked constituents; and

matching the search terms in the query against the entries in the index.

22. The method of claim 20 wherein the first textual input comprises a document and the second textual input comprises an information retrieval query and wherein preferentially matching comprises:

obtaining an index having entries corresponding to the document, the entries corresponding to the higher ranked constituents having higher weighting values associated therewith and the entries corresponding to the lower ranked constituents having lower weighting values associated therewith; and

matching the search terms in the query against the entries in the index based on the higher and lower weighting values.

23. A method of determining a relationship between first and second textual inputs, the method comprising:

analyzing the first textual input to obtain relations of constituents thereof;

determining a relative importance of the constituents in determining the relationship between the first and second textual inputs based on the relations obtained; and

determining the relationship between the first and second textual inputs based on the constituents and the relative importance of the constituents.

24. The method of claim 23 and further comprising:

obtaining a threshold importance and wherein determining the relationship comprises determining the relationship based only on constituents having a relative importance above the threshold importance.

25. The method of claim 23 and further comprising:

assigning a weighting value to each constituent based on the relative importance corresponding to each constituent, and wherein determining the relationship comprises determining the relationship based on the weighting values associated with the

constituents.

26. The method of claim 23 wherein the first textual input comprises a document and the second textual input comprises a query and wherein determining the relationship between the first and second textual inputs comprises determining similarity in meaning between the document and the query.

27. The method of claim 23 wherein the first and second textual inputs each comprise documents and wherein determining the relationship comprises determining a similarity in meaning between the documents.

28. The method of claim 27 wherein determining the relationship further comprises determining whether the first and second textual inputs are to be clustered in a logical cluster based on the similarity in meaning between the first and second documents.

29. The method of claim 23 wherein the analyzing step comprises:

obtaining case information corresponding to the constituents in the first textual input.

30. The method of claim 29 wherein determining a relative importance comprises:

determining a relative importance of the constituents based on the corresponding case information.

31. The method of claim 23 wherein the analyzing step comprises:

obtaining grammatical relations corresponding to the constituents in the first textual input.

32. The method of claim 31 wherein determining a relative importance comprises:

determining a relative importance of the constituents based on the corresponding grammatical relations.

33. A computer readable medium storing an index of textual material used for determining a relationship between first and second textual inputs, the index comprising a data structure including:

a plurality of constituents from the textual material, the plurality of constituents having a predetermined usefulness in determining the relationship based on relations of constituents in the textual material.

34. The computer readable medium of claim 33 wherein the textual material includes a plurality of sentences, and wherein the index comprises:

a portion of a syntactic structure corresponding to each of the plurality of sentences, each syntactic structure being indicative of the grammatical relations of constituents in the corresponding sentences.

35. The computer readable medium of claim 33 wherein the predetermined usefulness corresponds to a location

of grammatical relations corresponding to each of the plurality of constituents on a predetermined hierarchy.

36. A computer readable medium storing an index of textual material used for determining a relationship between first and second textual inputs, the index comprising a data structure including:

a plurality of constituents from the textual material, the plurality of constituents having a predetermined usefulness in determining the relationship based on case information corresponding to the constituents in the textual material.

37. The computer readable medium of claim 36 wherein the predetermined usefulness corresponds to a location of the case information corresponding to each of the plurality of constituents on a predetermined hierarchy.

38. A computer readable medium storing an index of textual material used for determining a relationship between first and second textual inputs, the index comprising a data structure including:

a plurality of constituents from the textual material, the plurality of constituents having predetermined characteristics indicative of usefulness in determining the relationship and being annotated with annotations indicative of the usefulness.

39. The computer readable medium of claim 38 wherein

the constituents are annotated with a binary annotation indicating whether the constituents have corresponding grammatical relations located above a hierarchy threshold on a predetermined grammatical relation hierarchy.

40. The computer readable medium of claim 38 wherein the constituents are annotated with a weight value indicative of the usefulness of the constituents in determining the relationship.

41. A computer readable medium storing a data structure used in determining a relationship between first and second textual inputs, the data structure comprising:

a plurality of pre-computed aspects of at least one of the first and second textual inputs, the pre-computed aspects being useful in determining the relationship between the first and second textual inputs.

42. The computer readable medium of claim 41 wherein the plurality of pre-computed aspects include:

a linguistic analysis of at least a portion of the first or second textual input.

43. The computer readable input of claim 41 wherein the plurality of pre-computed aspects include:

a plurality of constituents of the first or second textual input; and

a predetermined indication of usefulness, associated with the plurality of constituents, in determining the

relationship between the first and second textual inputs.

44. A method of generating an index corresponding to a first textual corpus, the index for use in determining a relationship between the first textual input and a second textual input, the method comprising:

identifying first constituents in the first textual input having relations indicative of usefulness in determining the relationship; and

generating the index based on the constituents identified.

45. A method of identifying a hierarchy threshold in a relation hierarchy corresponding to a textual input, wherein the relation hierarchy locates relations therealong according to usefulness in determining a meaning of the textual input, the method comprising:

(a) selecting a hierarchy threshold;

(b) identifying constituents in the textual input having corresponding relations below the hierarchy threshold;

(c) obtaining an indication of the usefulness of the constituents, other than the identified constituents, in determining the meaning of the textual input; and

(d) determining whether the selected hierarchy threshold is a suitable threshold based on the indication of usefulness.

46. The method of claim 45 wherein the obtaining

step (c) comprises:(c)(1) removing from the textual input all constituents, having a corresponding relation located on the relation hierarchy below the selected hierarchy threshold;

(c)(2) providing an index corresponding to the textual input with the constituents removed;

(c)(3) performing information retrieval operations on the index;

(c)(4) determining a reduction in a size of the index achieved by removing the constituents; and

(c)(5) determining performance of the information retrieval operations.

47. The method of claim 46 wherein the determining step (d) comprises:

(d)(1) if performance is adequate, given the reduction in the size of the index, identifying the selected hierarchy threshold as a suitable threshold.

48. The method of claim 46 wherein the determining performance step (c)(5) comprises:

determining at least one of speed and accuracy of the information retrieval operations.

49. The method of claim 45 and further comprising:

repeating steps (a) - (d) until a highest suitable threshold has been selected.

50. The method of claim 45 and further comprising:

if the hierarchy threshold is unsuitable,

conducting failure analysis to determine whether the selected hierarchy threshold should be conditionally identified as a suitable threshold.

51. The method of claim 45 and further comprising performing steps (a) to (d) in the order listed.

52. A method of identifying case information corresponding to constituents in a textual input, wherein the case information is indicative of usefulness in determining a meaning of the textual input, the method comprising:

(a) selecting first case information;

(b) identifying constituents in the textual input corresponding to the first case information;

(c) obtaining an indication of the usefulness of the identified constituents in determining the meaning of the textual input; and

(d) determining whether to identify the first case information as being useful in determining the meaning of the first textual input based on the indication of usefulness.

53. The method of claim 52 wherein the obtaining step (c) comprises:

(c)(1) removing from the textual input all constituents, having a case indicated by the first case information;

(c)(2) providing an index corresponding to the textual input with the constituents removed;

(c)(3) performing information retrieval

operations on the index;

(c)(4) determining a reduction in a size of the index achieved by removing the constituents; and

(c)(5) determining performance of the information retrieval operations.

54. The method of claim 53 wherein the determining step (d) comprises:

(d)(1) if performance is adequate, given the reduction in the size of the index, identifying the selected case information.

55. The method of claim 52 and further comprising:

if the selected case information is not useful, conducting failure analysis to determine whether the selected case information should be conditionally identified as useful.

56. The method of claim 52 and further comprising:

performing steps (a)-(d) a plurality of times; and

placing the selected case information in a ranked list based on the corresponding indication of usefulness.

57. A method of obtaining a relation hierarchy for use in determining a usefulness of constituents in determining a meaning of a textual input, the method comprising:

(a) selecting a first relation;

(b) identifying constituents in a textual corpus having the selected relation; and

(c) obtaining an indication of the usefulness of the identified constituents in determining the meaning of the textual input;

(d) repeating steps (a) - (c) for each desired relation; and

(e) ranking the desired relations in the relation hierarchy based on the indication of usefulness for each desired relation.

58. The method of claim 57 wherein the obtaining step (c) comprises:

(c)(1) indexing constituents in a textual corpus having the first relation;

(c)(2) performing information retrieval operations on the index; and

(c)(3) determining performance of the information retrieval operation.

59. The method of claim 58 wherein the ranking step (e) comprises:

(e)(1) ranking desired relations having higher performance higher in the relation hierarchy.

60. The method of claim 57 wherein selecting a first relation comprises:

selecting a first grammatical relation.

61. The method of claim 57 wherein selecting a first relation comprises:

selecting a first case relation.

SUB C47 62. A method of determining a relationship between

first and second textual inputs, the method comprising:

obtaining a hierarchy of relations;

obtaining a hierarchy threshold based on a usefulness of the relations in the hierarchy in determining the relationship between the first and second textual inputs;

identifying constituents in the first textual input that have the relations in the hierarchy;

determining a usefulness of the identified constituents by locating the identified constituents in the hierarchy; and

determining the relationship between the first and second textual inputs based on identified constituents having associated relations above the hierarchy threshold.

add A7